Int. J. Production Economics 147 (2014) 634–646

Contents lists available at ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Investigating the impact of operational variables on manufacturing cost by simulation optimization

Rui Zhang [a], Wen-Chyuan Chiang [b,*], Cheng Wu [c]

[a] School of Economics and Management, Nanchang University, Nanchang 330031, PR China
[b] Collins College of Business, The University of Tulsa, Tulsa, OK 74104, USA
[c] Department of Automation, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 28 December 2010
Accepted 10 April 2013
Available online 18 April 2013

Keywords:
Operational variable
Manufacturing cost
Simulation optimization
Return on investment
Machine scheduling
Particle swarm optimization

ABSTRACT

In this paper, we focus on the relationship between operations-based variables (specifically, production speed, scrap rate and maintenance speed) and the manufacturing cost. These variables usually produce opposite influences on the variable cost and the fixed cost. For example, setting the production speed at a high level is beneficial for reducing the variable cost. However, maintaining the high speed incurs considerable fixed costs at the same time. Therefore, an optimization approach is necessary to determine the optimal values of the operational variables for minimizing the average cost. First, a discrete-event simulation procedure is designed for describing the stochastic production environment and for evaluating the settings. Then, an optimization approach based on the ordinal optimization (OO) philosophy and particle swarm optimization (PSO) is used to search in the continuous space of the operational variables. In this process, the optimal computing budget allocation technique is applied so as to fully utilize the computational resource and potentially save the computational time. Finally, numeric computations are conducted for verifying the effectiveness of the proposed algorithm. Sensitivity analysis and discussions are also presented.

1. Introduction

In the production and operations management area, traditional research has focused on the productivity and efficiency of manufacturing systems. For example, makespan (Melouk et al., 2004; Damodaran et al., 2006) or total tardiness (Schaller, 2007; Biskup et al., 2008) is commonly adopted as the objective function in the studies of production scheduling. Makespan (often noted as $C_{\max}$) is also known as the ultimate completion time of a set of jobs and thus is an index for describing the productivity (or capacity) of the production system. Total tardiness ($\sum T_i$) is an index for describing the ability to meet the requirements of customers (i.e. service quality) and thus reflects the firm's management efficiency. Of course, these two types of performance measures are able to capture some key issues of concern for operations managers. However, it is gradually recognized since the late 1990s that operations management must interact with finance (Cachon and Terwiesch, 2009, Chapter 5) and other areas in order to achieve profitability. For example, profitability analysis is an important aspect in the research of corporate finance. Thus, it is interesting to study the relationship between operational performance and profitability (Soteriou and Zenios, 1999), as well as how to optimize the organization of operational settings for maximizing the profit of a manufacturing firm.

In the literature, there are several publications which investigate the link between operational variables and financial performance in service organizations. In particular, the US airline industry has been widely examined as a subject of study. For example, Tsikriktsis (2007) uses regression method to explore the historical data of major US airline companies (released by the US department of transportation) and finds that some operational variables (e.g. capacity utilization) are clearly linked with profitability. Lapré and Scudder (2004) present a model for describing the relationship between service variables (e.g. lost luggage, overbooking) and the financial performance. However, the similar research is scarce with regard to the manufacturing industry, partly because the manufacturing companies have not been forced to disclose the operational data to the public (as required in the airline industry). Therefore, in order to perform such a research for the manufacturing sector, simulation is a viable approach. The present paper is just an attempt in this area.

In most existing research for production systems, random events have not been fully considered. For instance, a usual assumption in production scheduling research (Haouari and Hidri, 2008; Wang et al., 2008) is that the machines are continuously available and no

* Corresponding author.
E-mail addresses: r.zhang@ymail.com (R. Zhang), wen-chiang@utulsa.edu, wen-chyuan-chiang@utulsa.edu (W.-C. Chiang), wuc@tsinghua.edu.cn (C. Wu).

0925-5273/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2013.04.018

defects will be produced. This is of course inconsistent with the real-world situation. Indeed, if random factors (such as machine breakdown and job failure) are incorporated into the research model, we will face a discrete event dynamic system (DEDS) (Cassandras and Lafortune, 2008), and thus discrete-event simulation (Banks et al., 2009) is needed for the performance evaluation of such systems involving randomness. Therefore, in this paper, we adopt simulation-based optimization to find satisfactory settings for the adjustable operational variables in the manufacturing system. The considered objective to minimize is the unit manufacturing cost, which is a measure for profitability if the price of the product is determined exogenously (e.g. according to the market conditions).

The rest of the paper is organized as follows. Section 2 introduces some topics related with our research and methods. Section 3 defines the model of the problem we are to study. Section 4 describes the discrete-event simulation procedure for the discussed production model. Section 5 proposes a crude model for fast evaluation of operational settings to the production system. Section 6 describes the approach we propose for optimizing the operational variables in the system, which is based on the ordinal optimization methodology and particle swarm optimization. Section 7 introduces the numerical setup for the computational experiments. Section 8 presents the key computational results and comparisons. Section 9 performs sensitivity analysis on the long-term parameters in the system, leading to a discussion about the return on investment. Finally, Section 10 concludes the paper.

2. Related research background

2.1. Uniform parallel machine scheduling

In the parallel machine scheduling problem, we consider n jobs that are waiting for processing. Each job consists of only a single operation which can be processed on any one of the m machines $M_1, \ldots, M_m$. As a conventional constraint in scheduling models, each machine can process at most one job at a time, and each job may be processed by at most one machine at a time.

There are three types of parallel machines (Brucker, 2007):

- Identical parallel machines (denoted by "P"). The processing time p_j^k of job j on machine k is identical for each machine, i.e. $p_j^k = p_j$.
- Uniform parallel machines (denoted by "Q"). The processing time p_j^k of job j on machine k is $p_j^k = p_j/q_k$, where q_k is the speed of machine k.
- Unrelated parallel machines (denoted by "R"). The processing time p_j^k of job j on machine k is $p_j^k = p_j/q_{k,j}$, where $q_{k,j}$ is a job-dependent speed of machine k.

If preemption of operations is not allowed, scheduling uniform parallel machines with the makespan (maximum completion time) criterion (described as $Q||C_{\max}$) is a strongly NP-hard problem. According to the reduction relations between objective functions (Pinedo, 2008), scheduling uniform parallel machines under due-date-related criteria (e.g. total tardiness) is also strongly NP-hard. Therefore, meta-heuristics have been widely used for these problems (Raja et al., 2008).

2.2. Simulation optimization and ordinal optimization (OO)

The simulation optimization problem is generally defined as: find a setting of parameters (θ) which minimizes a given objective function $J(\theta)$, i.e.

$$\min_{\theta \in \Theta} J(\theta),$$

where Θ represents the search space for the optimization variable θ. The key assumption in simulation optimization is that $J(\theta)$ is not directly available, and simulation is the only way to acquire the evaluation of θ. Moreover, since practical simulation procedures must make a trade-off between accuracy (which intends to include as many details as possible) and time performance (which demands the simulation be as fast as possible), simulation can only provide a noisy estimation of $J(\theta)$, which is usually noted as $\hat{J}(\theta)$.

Simulation optimization problems often arise in the design of complex systems such as electric power grid, large-scale computer network, and semiconductor manufacturing system. The variable θ corresponds to the controllable parameters in the system to be designed (for this reason, θ is usually called a setting rather than a solution in terms of simulation optimization). In practice, it is unaffordable to perform the parameter tuning process on the real system itself, so simulation optimization is useful for finding a good enough setting for the system without expensive experiments. However, three difficulties naturally exist in the implementation of simulation optimization: (1) the search space (Θ) is often very large, containing zillions of choices for the system parameters; (2) simulation is usually very time-consuming for real-life systems; (3) simulation is subject to random errors, so a large number of simulation replications have to be adopted to get the correct evaluation of θ.

These problems all suggest that simulation optimization can be extremely costly in terms of computational burden. For a survey of existing methods for simulation optimization, interested readers can refer to Fu and Glover (2005). Here we will focus on the ordinal optimization (OO) methodology, which was first proposed by Ho et al. (1992).

OO attempts to settle the above difficulties by emphasizing two important ideas: (1) order is much more robust against noise than value; (2) aiming at the single best solution is computationally expensive, so it is wiser to focus on the "good enough". It not just mentions these ideas in words, but the major contribution of OO is that it quantifies these ideas and thus OO can provide accurate guidance for our optimization practice. We will list the main procedure of basic OO as follows. Meanwhile, we strongly suggest interested readers to turn to Ho et al. (2007) for detailed theory and proofs.

Suppose we want to find k settings that belong to the top-g (normally $k < g$). Then, basic OO consists of the following steps:

Step 1: Uniformly and randomly select N settings from Θ (this set of initial solutions is denoted by I).

Step 2: Use a crude and computationally fast model for the studied problem to estimate the performance of the N settings in I.

Step 3: Pick the observed top s settings of I (as estimated by the crude model) to form the selected subset S.

Step 4: Evaluate all the s settings in S using the exact simulation model, and then output the top k ($1 \le k < s$) settings.

As an example, let $g = 50$ and $k = 1$. If we take $N = 1000$ in Step 1 and the crude model in Step 2 has a moderate noise level, then OO theory ensures that the top setting in S (with $s \cong 30$) is among the actual top-50 of the N settings with probability no less than 0.95. In practice, s is determined as a function of g and k, i.e. $s = Z(g, k; N, \text{OPCclass}, \text{noiselevel})$.

The ordered performance curve (OPC) is a conceptual plot of the objective values as a function of the order of performance (i.e. the best, the 2nd best, and so on). It describes the solution distribution (and thus the difficulty level) of the considered problem. Noise level is used to describe the degree of accuracy of the crude model. Since $J_{\text{crude\_model}} = J_{\text{complex\_simulation\_model}} + \text{noise}$, the noise level can be